7th International Conference on

CLINICAL TRIALS

&

12TH WORLD CADD & DRUG DELIVERY SUMMIT

September 24-26, 2018 | Chicago, USA

Bioassay protocols: Semantic annotation to enable informatics

Luke S Fisher, Alex M Clark, Janice E Kranz, Hande Kucuk, and Barry A Bunin
Collaborative Drug Discovery, Inc,USA

Bioassay protocols are conspicuously absent from the informatics of drug discovery: Current best practices have not progressed beyond using scientific English text, which is intractable to software. We will present our solution which draws from the rich semantic web vocabularies of the BioAssay Ontology, Drug Target Ontology, Gene Ontology, and others. On their own, these ontologies are not friendly to experimental scientists, and so we have created the Common Assay Template, which turns the massive hierarchies of the underlying ontologies into useful guidelines. This has been supplemented by machine learning infrastructure to help translate existing text into suggestions, with the help of natural language analysis. Using our new web-based interface, a small team of biologists was able to annotate 3500 MLPCN screening assays that were extracted from the PubChem database, which consumed approximately 3 weeks FTE. These semantically annotated protocols are fully machine-readable, which imparts many new capabilities, which apply at all scales. Searching can be done using precise specific terms, which is far more effective than keyword searching. In conjunction with electronic lab notebooks, the annotations serve as a facile way to classify and organize experiments and keep tabs on the activities of colleagues. These well-defined annotations can also serve as an alternative to long-winded text. Applied to a large scale, the machine readability of these annotations enables a diverse array of algorithms to be applied to assay databases, such as clustering and selection of groups of compatible assays for model building, or analysis of the protocol designs and their effects on structure-activity relationships. Large numbers of annotated assay protocols from open data such as PubChem and ChEMBL are available for novel analyses for big pharma, biotechnology companies, academics, and consortia. We have explored a number of techniques for utilizing large quantities of data, including the development of searching interfaces, visualization modes, and methods for extracting related data and creating models. We have also studied specific trends within public screening data which we have elucidated with the help of our own curated content, and investigated some of the characteristics of specific projects, such as the NIH molecular libraries probes, which can be analyzed retrospectively given the much larger amount of information that is now available.

Biography

Luke S Fisher brings twenty years of experience in scientific informatics solutions. Managing Pre-Sales, Post-Sales and working in Account Management has expanded my domain knowledge of scientific informatics and provided me the ability to maintain a successful track record. He have covered a broad range of clients including the world's leading pharmaceutical, biotech, agricultural, chemicals, academic and government labs. He have extensive experience in scientific software solutions from the smaller scale deployment of point solutions like molecular modeling packages to the larger enterprise scale of ELNs, scientific workflow technologies, data content, analysis and visualization. It also includes managing the support complexity of software integration strategies based on numerous mergers and acquisitions..

Notes: